REMARKS

SECTION 103 REJECTIONS

Claims 1, 2, 5, 6, 9 and 11

In the Office Action, claims 1, 2, 5, 6, 9 and 11 were rejected under 35 U.S.C. §103(a) as being unpatentable over the admitted prior art (hereinafter, APA) in view of Frey et al. (U.S. Patent Publication 2002/0173953, hereinafter Frey) and in further view of Zangi et al. (U.S. Patent Publication 2004/0111258, hereinafter Zangi).

Independent claim 1 provides a method of determining an estimate for a noise-reduced value representing a portion of a noise-reduced speech signal. The method includes generating an alternative sensor signal using an alternative sensor other than an air conduction microphone and converting the alternative sensor signal into at least one alternative sensor vector in the cepstral domain. A weighted sum of a plurality of correction vectors is added to the alternative sensor vector to form the estimate for the noise-reduced value in the cepstral domain. Each correction vector corresponds to a mixture component and each weight applied to a correction vector is based on the probability of the correction vector's mixture component given the alternative sensor vector. An air conduction microphone signal is also generated and the air conduction microphone signal is converted into an air conduction vector in the power spectrum domain. A noise value is estimated and the noise value is subtracted from the air conduction vector to form an air conduction estimate in the power spectrum domain. The estimate of the noise-reduced value is converted from the cepstral domain to the power spectrum domain. The air conduction estimate and the estimate for the noise-reduced value are combined in the power spectrum domain to form a refined estimate for the noise-reduced value in the power spectrum domain.

Claim 1 is not shown or suggested in the combination of cited art. In particular, none of the cited art shows or suggests forming an estimate of a noise-reduced value in the cepstral domain, converting the estimate of the noise-reduced value from the cepstral domain to a power spectrum domain, and then combining the estimate of the noise-reduced value in the power spectrum domain with an air conduction estimate.

In the Office Action, it was asserted that Frey suggests determining a noise-reduced value in the cepstral domain and converting the estimate back to the power spectrum domain. Applicants respectfully dispute this assertion.

Frey does not convert a cepstral domain noise-reduced value back to the power spectrum domain. As shown in paragraph 40 of Frey, the cepstral domain values are used directly for speech recognition decoding. The fact that Frey does not actually show converting cepstral domain noise-reduced values to the power spectrum domain appears to be acknowledge by the Examiner because the Examiner goes on to say that "it would be obvious that the signal would have to be converted back to the spectrum domain in order for it to be used to represent the signal in a meaningful way, as cepstral analysis is a log scale." However, this statement is simply not true. During speech recognition, which Frey performs, the cepstral-domain representation provides a meaningful representation of the speech signal and is preferred, as indicated by Frey, when decoding speech. In fact, if one wanted to work in the power spectrum domain, there would be no reason to form the cepstral vectors in Frey. Instead, the noise-reduced values could be formed in the spectral domain instead. This would greatly simplify the calculations in Frey. However, Frey wants cepstral domain values and as such, it would not be obvious to convert the cepstral domain noise-reduced value of Frey back to the power spectrum domain.

Further, to form the combination suggest by the Examiner, one of the values in Zangi would have to be computed in the cepstral domain and then converted into the power spectrum domain while calculating the other values in the power spectrum domain. However, this may cause artifacts to arise since the transfer function for filters 74A-74M is based on the assumption that all of the transfer functions operate in the same domain. (see the equation in paragraph [0103]). In addition, it is easier to set all of the transfer functions in the same domain as shown by Zangi. As such, there would be no motivation to change Zangi as suggested by the Examiner.

Since none of the cited art shows or suggests forming an estimate of a noise reduced value in the cepstral domain, converting the noise reduced value from the cepstral

domain to the power spectrum domain and combining the noise reduced value in the power spectrum domain with an air conduction estimate in the power spectrum domain to form a noise reduced value in the power spectrum domain, claim 1 and claims 2, 5, 6, 9 and 11, which depend therefrom, are patentable over the cited art.

CLAIMS 12 AND 13

Claims 12 and 13 were rejected under 35 U.S.C. §103(a) as being unpatentable over Park et al. (U.S. Patent 5,590,241, hereinafter Park) in view of the APA and in further view of Griffin et al. (U.S. Patent 5,701,390, hereinafter Griffin).

Claim 12 provides a method of determining an estimate of a clean speech value. The method includes receiving an alternative sensor signal from a sensor other than an air conduction microphone and receiving an air conduction microphone signal from an air conduction microphone. A pitch frequency is identified for a speech signal based on the alternative sensor signal by identifying which frequency of a group of candidate frequencies is the pitch frequency. The pitch frequency is used to decompose the air conduction microphone signal into a harmonic component and a residual component by modeling the harmonic component as a sum of sinusoids that are harmonically related to pitch. The harmonic component and the residual component are used to estimate the clean speech value by determining a weighted sum of the harmonic component and the residual component.

Claim 12 is not shown or suggested in the combination of cited art because none of the cited art identifies which frequency of a group of frequencies is a pitch frequency for a speech signal based on an alternative sensor signal and none of the cited art determines an estimate of a clean speech value by determining a weighted sum of a harmonic component and a residual component where the clean speech value represents a noise reduced signal having a reduced noise relative to the noisy air conduction microphone signal.

In the Office Action, Park was asserted as showing the step of identifying a pitch for a speech signal based on an alternative sensor signal at column 3, line 21 because it produces a signal that has primarily low-frequency speech components. However, the cited section makes

no mention of identifying which frequency of a group of candidate frequencies is a pitch frequency for a speech signal. Simply producing an alternative sensor signal that has low-frequency speech components is not the same as identifying which of those low-frequency speech components is a pitch frequency for a speech signal.

In the Office Action, the Examiner addressed this argument by stating that since Applicant had not defined the group of candidate frequencies, the group could include all frequencies. However, even if the group of candidate frequencies is taken to be all possible frequencies, Park still does not show identifying which frequency of all possible frequencies is a pitch frequency for a speech signal based on an alternative sensor signal. Instead, it simply produces a signal with low-frequency speech components. Producing such a signal does not involve identifying which frequency is a pitch frequency for a speech signal and is not equivalent to identifying which frequency is a pitch frequency.

In addition, none of the cited references show or suggest estimating a clean speech value representing a noise reduced signal having reduced noise relative to the noisy air conduction microphone signal by determining a weighted sum of a harmonic component and a residual component.

In the Office Action, it was asserted that Griffin teaches forming a weighted sum of a harmonic component and a residual component in FIG. 2 where the voiced synthesis and unvoiced synthesis components are added to produce an estimated speech signal. However, the summation performed in FIG. 2 of Griffin is designed to reproduce the input signal. The entire goal of Griffin is to reproduce the input signal without distorting it anymore than it was already distorted when it is was input. There is no mention in Griffin of forming a clean speech value representing a noise reduced signal having reduced noise relative to a noisy air conduction microphone signal by determining a weighted sum of a harmonic component and a residual component.

In the Office Action, the Examiner addressed this argument by stating that it was moot since Griffin was only cited to show the formation of a signal using a combination of sinusoid frequency information and residual information. However, if Griffin is not being cited

to show the step of estimating a clean speech value representing a noise-reduced signal by determining a weighted sum of a harmonic component and a residual component, then the Office Action has not cited any reference that shows this limitation. Instead, the Office Action has cited APA that shows spectral subtraction of noise from a noisy speech signal to form a clean speech signal and Griffin that shows the formation of a noisy signal by adding voiced and unvoiced components together. There is no suggestion in either reference that adding together the voiced and unvoiced components of Griffin would produce a noise-reduced value. Instead, combining APA with Griffin would result at most in performing spectral subtraction after adding the voiced and unvoiced components together. That is substantially different from claim 12 where a clean speech value is formed by determining a weighted sum of a harmonic component and a residual component.

Further, Griffin does not even form a signal by determining a weighted sum of a harmonic component and a residual component. Although FIG. 2 of Griffin shows an unvoiced component and a voiced component being added together, it does not indicate that this addition is a weighted sum. It appears to be a simple sum, with no weights applied to the voiced and unvoiced portions.

In response to this argument, the Office Action cited column 13, line 62 - column 14, line 7 of Griffin where Griffin discusses forming the unvoiced portion of a signal using a weighted overlap and add. The Office Action further states that "the result is the residual component is weighed in the addition to the frequency components." Applicants respectfully dispute this assertion.

The assertion appears to be that since the overlap and add uses weighting, the resulting unvoiced portion is somehow weighted relative to the voiced portion in Griffin. However, Griffin does not suggest this is true. There are no other statements in Griffin that the unvoiced portion should be weighted relative to the voiced portion of the signal. As shown in Griffin, the signal that Griffin defines as the unvoiced component is added directly to a voiced component to form an output signal without weighting. The fact that a weighted overlap and add and forward and inverse FFT filtering are used to form the unvoiced component does not suggest

that the unvoiced component is being weighted relative to the voiced component. As such, Griffin does not form a signal using a weighted sum but instead forms a signal by determining the simple sum of a voiced component and an unvoiced component. This simple sum is clearly indicated in FIG. 2 of Griffin where the sum is shown as an unvoiced component $s_{uv}(n)$ being added to a voiced component $s_v(n)$. There are no weighting values in front of either term.

Since the cited art does not show or suggest identifying which frequency of a group of candidate frequencies is a pitch frequency based on an alternative sensor signal and because the cited art does not show or suggest estimating a clean speech value by determining a weighted sum of a harmonic component and a residual component, the combination of cited art does not show or suggest the invention of claim 12 or claim 13 which depends therefrom.

CLAIMS 14, 15, 17, 18, 23, 24, AND 29

Claims 14, 17, 18, 23, 24 and 29 were rejected under 35 U.S.C. §103(a) as being unpatentable over Park in view of Zangi and further in view of Frey. Claim 15 was rejected under 35 U.S.C. §103(a) as being unpatentable over Park in view of Zangi and in further view of APA.

Claim 14 provides a computer-readable storage medium storing computer-executable instructions for performing steps. The steps include receiving an alternative sensor signal from an alternative sensor that is not an air conduction microphone, receiving a noisy test signal from an air conduction microphone and generating a noise model from the noisy test signal. The noise model includes a mean and a covariance. The noisy test signal is converted into at least one noisy test vector and the mean of the noise model is subtracted from the noisy test vector to form a difference. An alternative sensor vector is formed from the alternative sensor signal. A correction vector is added to the alternative sensor vector to form an alternative sensor estimate of a clean speech value. A weighted sum of the difference and the alternative sensor estimate is set as an estimate of the clean speech value, wherein the weighted sum is computed using the covariance of the noise model to compute weights for the weighted sum.

Claim 14 is not shown or suggested in any of the cited art. In particular, none of the cited art shows or suggests a weighted sum that is computed using the covariance of a noise model to compute weights for the weighted sum. In the Office Action, Frey was cited as showing a noise model with a covariance. In addition, it was asserted that it would be obvious to consider the covariance when weighting as the covariance indicates how correlated the noise signals are, indicating the depth of the noise that is being filtered out. Applicants respectfully dispute this assertion.

There is no suggestion in the cited art that a measure of the "depth of noise that is being filtered out" should be used to determine weights for a weighted sum of a value formed by subtracting a noise mean from an air conduction microphone signal and an alternative sensor estimate of a clean speech value. The "depth of noise that is being filtered out" would not appear to be relevant to which of these two components should be more heavily weighted. There is no suggestion in any of the references that the correlation between noise signals used to train a noise model should be used to weight a clean speech value determined using a mean of the noise model any differently than a clean speech value determined from an alternative sensor signal. In fact, Applicants submit that it would never have occurred to the Examiner to use the covariance of a noise model to weight clean speech estimates if the Examiner had not first read Applicants' claim.

Since none of the cited references show or suggest setting a weighted sum as an estimate of a clean speech value where the weighted sum is computed using the covariance of a noise model to compute weights for the weighted sum, the combinations of cited art do not show or suggest the invention of claim 14 or claims 15, 17, 18, 23, 24 and 29, which depend therefrom.

CONCLUSION

In light of the above remarks, claims 1, 2, 5, 6, 9, 11-15, 17, 18, 23, 24 and 29 are in form for allowance. Reconsideration and allowance of the claims is respectfully requested.

The Director is authorized to charge any fee deficiency required by this paper or credit any overpayment to Deposit Account No. 23-1123.

Respectfully submitted,

WESTMAN, CHAMPLIN & KELLY, P.A.

By: ______________________________
Theodore M. Magee, Reg. No. 39,758
900 Second Avenue South, Suite 1400
Minneapolis, Minnesota 55402-3319
Phone: (612) 334-3222 Fax: (612) 334-3312

TMM:sew